DOCUMENT RESUME

ED 362 384 SE 053 654

AUTHOR Sobolewski, Stanley J.
TITLE A Latent Variable Path Analysis Model of Secondary Physics Enrollment in New York State.
PUB DATE Apr 93
NOTE 13p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (Atlanta, GA, April 17, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Classroom Research; *Enrollment Influences; High Schools; *Physics; *Science Education; Secondary School Science
IDENTIFIERS *New York

ABSTRACT

Physics is a fundamental science course and can be valuable to all students, however, enrollment at the high school level is low. This study tries to identify factors that influence high school student enrollment and achievement. Raw data was obtained from the New York State Education Department on magnetic tapes with over 300,000 records containing information about individual classes. From the data sets, variables determined to be appropriate for the model were selected, the latent variables identified, and their relationships explored. There are six major points identified in the study: (1) there has been little change in enrollment in the past 20 years; (2) school level variables, including size of school, instructional facilities, and teacher experience have the strongest relationship to physics enrollment; (3) variables external to the school have the biggest impact on physics achievement; (4) larger schools have disproportionately fewer students enrolled in physics; (5) schools with a higher percentage going to college have a greater physics enrollment; and (6) in schools with a higher percentage of females enrolled in physics, the relationship between achievement in previous science courses and physics achievement is much stronger. (PR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 362 384

A Latent Variable Path Analysis Model of Secondary Physics Enrollment in New York State

A presentation and discussion
at the Annual meeting of the
National Association of Research in Science Teaching
Atlanta, Georgia
April 17, 1993

by

Stanley J. Sobolewski
Department of Physics
Indiana University of Pennsylvania
Indiana, Pennsylvania 15705

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stanley J. Sobolewski

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SE 053 654

Introduction

The number of high school students choosing to enroll in physics has been traditionally low. In 1968, 20% of high school students in an average grade took a physics course (Welch 1968). A study conducted by the American Institute of Physics found that in 1987, 623,000 students were enrolled in a physics course. This represents about 20% of the 1987 high school graduates (Neuschatz and Covalt 1988). Since high school physics is the last science experience for most people, it is possible that up to 80% of the population has an incomplete science experience.

Physics is a fundamental science course and can be valuable to all students. Few educators would advise a student not to take some type of course in physics. Yet, the physics enrollments are low. The placement of the course at the end of the science curriculum sequence discourages some students from enrolling. In some cases, a lack of course availability may be a problem. According to the 1985-86 National Survey of Science and Mathematics Education, (Weiss 1987) when science teachers were asked to name a science course, only 12% of the teachers sampled mentioned physics. Other reasons for low enrollment in physics include students' perception that the course is difficult and the inaccessibility of the course due to prerequisites (Chandavarkar 1988, Gay 1978). Other explanations for the low physics enrollment include a lack of enthusiastic and qualified physics teachers (Franz, Aldridge & Clark, 1983).

Statement of Problem

This study will try to identify factors that influence high school physics enrollment and achievement.

Research Problem

The following broad research issues will be addressed:

1. What are some of the elements, both in the school environment and external to the formal educational process that have an influence on physics enrollment in high school? Included will be variables that have been discussed by other authors as well as new variables to be introduced in this study.

2. Will the model identify factors that differentiate schools with high PEP from schools with low PEP? Will the model identify factors that differentiate schools with high average scores on the Regents physics exams from schools with low average scores on the Regents physics exams?

Method

Data Acquisition and Analysis

The raw data was obtained from the New York State Education Department on five magnetic tapes. There were three sets of data on the tapes. The Basic Educational Data System(BEDS) tape held over 300,000 records that contained information about individual classes. Each record contained information on class size, the number of times the class meets per week, the quality of the students in the class (remedial to honors) as well as information on the teacher's experience and education. The school data tape is a compilation of surveys completed by the principal of each building. Data that pertains to the school's resources and students as a whole was found in this file. The data included the number and type of classrooms in the school, and the number of books in the library as well as the distribution of various minorities in the student body. Lastly, the Comprehensive Achievement Report (CAR) consisted of data on three magnetic tapes, which gave school level reports on the number of students taking and passing various state exams. Since school districts are required by the state to submit this information, there was little problem with missing data. The exception was with teacher's salaries and class size in small districts where physics and chemistry were taught in alternating years. The teacher's salary was not used as a predictor variable, since the salary is primarily a reflection of the number of years of experience in a particular district and therefore is redundant data. The missing physics enrollment for the 1990-1991 school year was replaced with one-half of the 1989-1990 physics enrollment, which was also included among the information on the magnetic tapes.

From the data sets, variables which were determined to be appropriate for the model were selected. These variables are listed in the appendix. From these data, the latent variables were subsequently identified. The previous work of Bryant (1979) as well as the models of Noonan and Wold (1983) was used as a guideline for the exploratory factor analysis. Initially, a factor analysis on all of the variables was conducted to determine the number of factors operating on physics enrollment. Secondly, groups of variables which were suspected of having a high correlation

with each factor were examined by the investigator to determine if these manifest variables reflected only one latent variable.

Model Construction

Essentially, the construction of the exploratory model was a three step process: 1) processing the data into a meaningful format 2) identification and construction of the latent variables and 3) determining the relationship between each of the variables. For each of these steps, a computer based procedure was used. For the first step, the data entry process in SPSS-X was used to read the raw data from the tapes. The second step used the procedure FACTOR, also an SPSS-X procedure. Thirdly, LISREL was used to verify the proposed model. Also, a series of multiple regressions was used to determine relationships between the manifest variables. The critical decision in the study is the development of latent variables. Once these variables have been identified and constructed, the relationship between the variables can be studied.

The process of factor analysis is a process of grouping variables which are highly correlated with each other. It is the working assumption of any factor analysis that there is a common element to the variables which are correlated with each other. The task of the researcher is to determine if the factors isolated by the analysis are representative of reality or spurious associations. The factor loading and coefficients were found using the SPSS-X statistics package. After several analysis had been performed, approximately eight factors emerged. These factors or latent variables were identified and named by the manifest variables associated with them. The factors identified and the rationale for postulating the analogous latent variable is as follows.

Initially the manifest variables listed in the appendix were used in an exploratory factor analysis. In this process it was found that some of the variables shard little variance with most of the other variables. The amount of shared variance is indicated by the communality (h^2) For example, the percent of students in a physical science class (NPHYSCI $h^2 = 0.07536$) or the number of black and white televisions in the school (BWTV $h^2 = 0.07413$) were not used since they had little common variance with most of the other variables. It is not assumed that these omitted variables are unimportant or that they have no predictive power; it is an attempt to focus the model on the problem of physics enrollments. When these weakly linked variables were removed, additional factor analysis were performed to isolate the latent variables.

Latent and Composite Variables

A latent variable is one which cannot be directly measured, while manifest variables can be measured. The nature of the latent variables must be determined by the theoretical framework and the relationship between the manifest variables. A latent variable is not constructed by the researcher, it is discovered in a way analogous to the discovery of an astronomical body. Unlike latent variables, composite variables are groups of measured variables rearranged by the experimenter. Composite variables make the process of identifying a simpler structure more efficient and reduces the amount of data necessary to construct the model. Composites also simplify the interpretation of the effect since fewer variables are easier to conceptualize. (Darmondy 1984) Among the consideration for identifying the best combination of variables to create the composite, Darmondy suggests the following:

a) the nature of the variables available

b) their relationship to the criterion variable

c) the researcher's preference for defining a latent pre-existing variable whose aspects are reflected in the components

d) the plausibility of the resulting construct

When appropriately chosen, the variables will maximize the variance explained in the criterion variable and produces a composite which conforms best to the method of analysis used. The addition of variables will increase the explained variance. Once the explained variance does not increase, there is no reason to increase the number of variables

A third relevant point made by Darmondy has to do with the selection of the specific variables. On the basis of a factor analysis, groups of variables should be selected that have a) high commonalties b) high factor loading c) positive correlation with the criterion and d) positive inter-correlation.

In most circumstances, if the composite variables are constructed appropriately, the latent variables in a model will agree with the composite variables in the same model.

The Path Analysis Diagram

The path analysis diagram essentially is the graphical representation of a set of simultaneous equations. There are many conventions on the construction of path diagrams, the method of representing associations is usually dependent upon the technique used to solve the system of equations. Here the conventions of Joreskog and Sorbom (1984) will be used since LISREL was used as the primary analytical tool.

LISREL - Linear Structural Relationship

Developed by Joreskog and Sorbom, LISREL 7 is a computer program that solves sets of linear equations. While the basic function of the program is not unique to LISREL, authors refer to sets of structural equations as a LISREL model. LISREL was used in this study primarily due to the nature of the data as well as convenience. The "full LISREL model" was used in this study. This full model consists of two types of equations. The first is referred to as the structural model and it refers to the relationship between the latent variables:

$$\eta = \beta(\eta) + \gamma(\xi) + \zeta \qquad \text{(i)}$$

The η vector contains elements that are referred to as the endogenous variables, or latent dependent variables. Vector ξ contains the exogenous variables, or latent independent variables. The ζ term is the error not explained by the model. β is a coefficient matrix that gives the influence of the η's on each other and γ is the matrix that contains terms that relate the endogenous and exogenous terms.

The second set of LISREL equations is referred to as the measurement model. These two equations are:

The process of identifying the latent variables and discovering the relationships between the latent variables and manifest variables was one of the primary tasks of this study. As mentioned above, a series of factor analyses were conducted to determine the number of latent variables or factors that operate in the schools of New York State.

$$y = (\Lambda_y)\,\eta + \varepsilon \qquad \text{(ii)}$$

$$x = (\Lambda_x)\,\xi + \delta \qquad \text{(iii)}$$

These relate the measured values (x and y's) to the latent variables (η and ξ). If each latent variable has a single indicator, then $\eta = y$ and $\xi = x$ and equation (i) becomes: $y = (B)y + (\Gamma)\,x + \zeta$ (i`)

Here, the B is the collection of all the β's and the Γ is the collection of all the γ's.

In the full LISREL model there are eight matrices that represent the relationships between: 1) the dependent manifest variables and the dependent latent variables 2) the independent manifest variables and the independent latent variables 3) the relationship between the dependent and independent latent variables 4) the relationship between the independent latent variables 5) the relationship between the dependent latent variables 6) the error, or variance in the manifest variables not explained by the model 7) the error in the latent variables and 8) the covariance matrix of the error terms of the latent variables.

Model 1 - Identification of Variables

The process of identifying latent variables requires the use of judgment as well as mathematical logic. The relationship that the variables have with each other is a function of their definition. Therefore, more than one model was developed using linear structural modeling. It is surmised that comparing and contrasting between the two models will increase practical understanding of science instruction in New York State Schools. In both cases, a general model was developed and confirmed with the LISREL computer program. Path coefficients that were found to be insignificant were eliminated. Path coefficients that had been omitted and were found to have a high and significant modification index were added to the model. The modification index is a non-scaled value that is assigned by LISREL to all of the path coefficients which are forced to be zero by the experimenter. Variables which are

highly related and have path coefficients fixed to zero are assigned a high modification index. This is an indication to the experimenter that the variables might be related. It is the decision of the experimenter to change the model and have the program calculate a value for a coefficient for a path which had been initially set to zero.

The initial structure developed for model 1 was influenced by the separation of variables into two categories, one of variables external to the control of the school and one external to school control. From the factor analysis illustrated below, eight factors were postulated, four exogenous and four endogenous.

Factor Analysis Identification of Latent Variables

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
PBIOPAST	.788				
PBIOENRT	.749				
PCS3PAST	.704				
PCS3ENRT	.692		.428		
PPHYPAST	.681				
PCHMPAST	.664				
COMMUNIT	.641				
PCHMENRT	.639		.515		
ATTENDAN	.639				-.588
PPHYENRT	.460		.420		
STUSECOM		.821			
NBOOKS		.750			
N4COLLEG		.740	.428		
NLABS		.618			
PPHYAP			.658		
PAPCAL			.649		
PCHEMAP			.631		
PAPBIO			.492		
TEAAGE				.871	
SCHEXPER				.843	
MEANDEG		.418		.590	
PDROPOUT					.843
PEARENRT	.420				
PEARPAST	.434				

	FACTOR 6	FACTOR 7
PEARENRT	.718	
PEARPAST	.717	
PPHYL		
PBIOG		.606
PCHEMG		.604
PBIOL		-.584
PCHEML		-.481

BEST COPY AVAILABLE

Eight latent variables were identified from the previous factor analysis. These latent variables were found to be the following:

Community Size (ξ-1) : The variable which is external to the school district, and is descriptive of the community. In this specific case, it is the size of the community.

Facilities (ξ-2) : Variables which are characteristic of the educational resources that are available for use at the school. This includes the number of science class rooms per student, the number of video-cassette players as well as the number of books per student in the school library.

Students Attendance (ξ-3) : Variables which describe the student body of the school, such as dropout rate and the average daily attendance.

Teacher Experience (ξ-4) : Variables which are characteristic of teachers' experience. These include the highest degree earned by a teacher as well as the number of years a teacher has taught in the district.

Regents Chemistry and Mathematics (η-1) : Variables which are descriptive of the "upper regents" classes, that is Regents chemistry and course III mathematics. This includes the proportion of student enrolled in the class as well as Regents test scores.

AP Enrollment & College Bound Graduates (η-2) : Variables which are characteristic of the college preparatory program at a school. This includes enrollment in advanced placement courses as well as the percent of graduates that attend a four year college.

Regents Biology and Earth Science (η-3) : Variables which are characteristic of the "lower regents" classes, such as earth science and biology.

Local Chemistry and Biology (η-4) : Variables which are characteristic of classes which do not have a state mandated curriculum. This includes enrollments in local biology and local chemistry.

In addition to these variables, there are two endogenous dependent latent variables, Physics Enrollment (η-5) and Physics Achievement (η-6). Each of these has only one manifest variable associated with it. These are, the percent of an average year enrolled in physics (PEP) and the percent of the students enrolled in Regents physics passing the Regents exam, respectively.

Since latent variables are not directly observed, they can in principle have arbitrary units of measure. However, since the purpose of constructing a path diagram is to determine the relationships between variables, units of measure must be assigned to the latent variables so that meaningful relationships can be found. Here the scales of measure for the latent variables will be the same as that of the manifest variable which accounts for most of the variance in the measure of the latent variable. The exception to this was in assigning one of the two primary independent variables, the percent of students passing the Regents chemistry exam as defining the scale for the **Regents** latent variable, even though the percent of students enrolled in Math course III accounted for more variance. It was felt that since physics is the primary subject under investigation, it would be a more appropriate scale.

The following table lists the latent variables, the manifest variable used to set the scale and the actual unit of measure used.

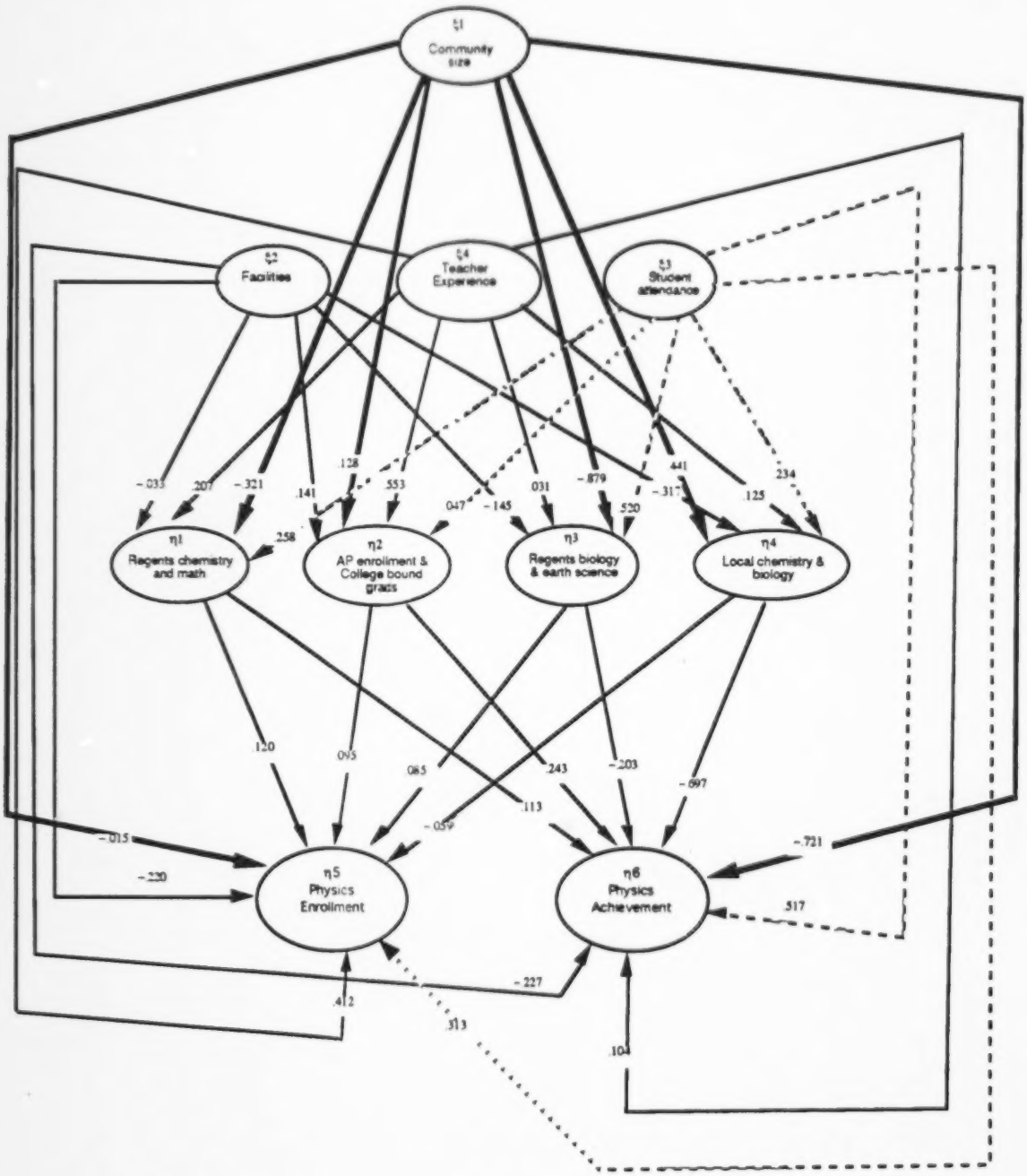
Latent Variable	Manifest Variable	Unit of Measure
Community	Size of community	6 point scale *Big Five* *Large City* *Small City* *Suburban* *Small town* *Rural*
Facilities	Number of books in library	Number
Students Attendance	Daily attendance	percent of students
Teachers	Teacher degree	7 point scale *Doctorate* *Masters plus* *Masters* *Bachelors plus* *Bachelors* *Normal School* *Secondary School*
Regents	Percent enrolled in physics	percent of students
AP Enrollment College	Percent of last years graduates attending 4 year colleges	percent of students
Regents Bio and Earth Science	Percent enrolled in Regents Biology or Earth Science	percent of students
Local Courses	Percent in Local Biology and Chemistry	percent of students

The latent variable model

After the latent variables have been defined by the collection of manifest variables, the relationship between the latent variables is calculated. In the terminology of LISREL, the values of β and γ are calculated. The β terms relate the endogenous latent variables to each other and the γ relate the exogenous latent variables to the endogenous latent variables. Paths that were postulated and were not statistically significant were omitted from the diagram. In the following diagram, the four exogenous latent variables are at the top of the page. The "Community size $\xi 1$" latent variable is at the top of the diagram for clarity of the diagram. In the logical structure of the diagram, it is at the same level as "Facilities" "Teacher Experience" and "Student Attendance." In general, however, there is a trend of general or "up-stream" variables at the top of the page and more specific "down stream" variables at the bottom of the page. At the bottom of the diagram are the two dependent latent variables, "physics enrollment" and "physics achievement." Each of these latent variables controls only one manifest variable: the percentage of students enrolled

in all physics classes for "physics enrollment" and the percent of Regents physics students who pass the New York State Regents physics exam for "physics achievement."

When the values of the path coefficients are first examined, it can be seen that there is a range of values, from 0.033 to 0.879. These path coefficients represent the change in standard scores of the dependent variable when the independent variable is changed one unit.



When the paths that lead to the two independent variables of physics achievement and physics enrollment are examined, A general trend in the difference between the paths leading to the enrollment and the achievement

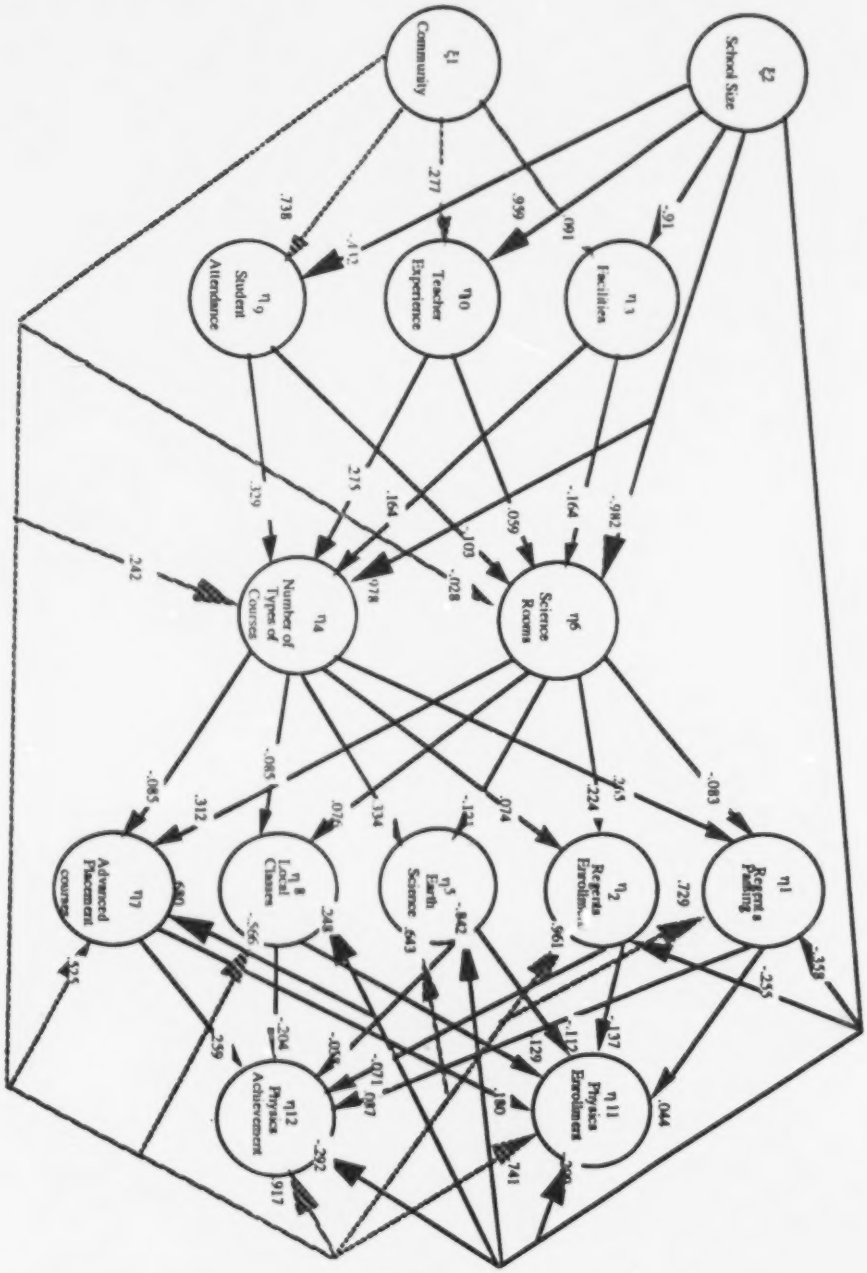
variable can be noted. The paths that lead to the physics enrollment latent variables are all small; teacher experience at .412, student attendance at .313, school facilities at -.220. All of the remaining variable path coefficients to the physics enrollment was .10 or less. The paths that lead to the physics achievement are slightly higher, the highest is the path from "community size" to physics achievement, with a value of -0.72. Local chemistry and biology are also are strongly associated with physics achievement (-.697) as well as student attendance (.517).

From this it can be concluded that the one of the strongest forces in increasing enrollment in physics is the experience of the teacher. While the teacher can bring the students into the classroom, the variables that are some what outside the control of the school, that is the size of the community and the student's attendance patterns, are more closely related to physics achievement. This strong relationship to the community latent variables can be also seen in the local chemistry and biology variable (.441), as well as the Regents biology and earth science variable (-.879). This negative path coefficient here indicates that the larger communities have a lower percentage of students enrolled in Regents biology and earth science. This trend of a strong relationship between community and enrollment is broken in the AP enrollment latent variables, where the experience of the teacher has the strongest relationship to AP enrollment. While it is possible that there is an association between the AP enrollment and teacher experience, it also is possible that the variables external to the school, such as the size of the community, is acting through the teacher experience variable. It is not possible to determine that potential relationship from this diagram; the community size and teacher experience (.553) are both exogenous latent variables and by design are independent of each other.

Essentially, the model illustrates the combined effects of school and community variables on physics enrollment and achievement. In general, schools that have high physics enrollment also have more experienced teachers. Schools that have a high passing rate on the Regents physics exam tend to be in smaller communities.

Model 2

In an attempt to better describe physics enrollment patterns, a second model of physics enrollment is presented here. The primary difference between the first and second model is in the number of manifest variables used to define the latent variables and the number of categories in which the latent variables were grouped. The second model has more of both. It is not necessarily true that one model is better than the other, since they both have statistically significant path coefficients and both have a high goodness of fit index (0.77 for model 2 and .83 for model 1). The motivation for the second model was the need for a level of intermediate school wide variables. The second model also has more manifest indicators for the latent variables. In the case of the first model, most of the latent variables had one or two indicators. In the case of the second model, most of the latent variables have from three to six manifest variables associated with each latent variable. This second model is illustrated here.



The latent variables in this model were organized in levels from the most global or general to those that are most specific. This allow the variance to flow either from the exogenous variables directly to the dependent latent variable or to flow through the school through school level variable, science department level variables, and physics instruction variables. The most striking result is that most of the variance in both physics enrollment and physics achievement can be associated with the variables which do not flow from the school. However, this finding is consistent with some other research than indicates similar amount of common variance between school variables and achievement: 30% Spelhaug (1990), 5-15% Talton (1983). This second model used 30 manifest variables and 14 latent variables to explain physics enrollment and physics achievement in schools in New York State. According to the recommendations of Bryant (1974) and others, the model has met the goal as described above. Specifically, 1) each of the variables in the model were theoretically justified 2) the model had several zero path coefficients 3) all of the correlations were statistically significant 4) all of the regression equations used to calculate the path coefficients were significant at the 0.01 level 5) none of the beta weights had a magnitude less than one standard error of the beta weight 6) At least 50% of the variance in the PEP is predicted by the model.

The model meeting the above criteria in combination with an adjusted goodness to fit index of 0.85 indicates that the model constructed here is a satisfactory model of physics enrollment and achievement in New York State high schools.

Results

There are six major points that can summarize the results of this study. 1) According to the variables that were used here, there has been little change in the enrollment of students in high school physics in New York State in the past 20 years. The average PEP in New York State is around 20%, which is a similar figure to what has been found in previous studies. The factors that influence the enrollment also were identify as being the same in 1991 as in 1971. 2) School level variables, which include the size of the school, the instructional facilities and teacher experience have the strongest relationship on physics enrollment. 3) Variables external to the school, such as community size and student ability have the greatest effect on physics achievement. 4) Larger schools have disproportionately fewer students enrolled in physics. 5) Schools with a higher percentage of graduate attending college also have a higher PEP They also have a greater percentage of teachers with advanced degrees. 6) In schools with a higher percentage of females enrolled in physics, the relationship between achievement in physics and achievement in previous science courses is much stronger.

The combination of results 2) and 3) indicate the independent nature of achievement and enrollment. The school is capable of increasing the enrollment of physics, but it is essentially the student and the student level variables that are more associated with success in physics that the school level variables.

Part of the purpose of this study was to investigate the development of educational models. Therefore, two latent variable structural models were constructed. In the comparison of the two models, some of the relationships between variables were similar. The strength of a relationship between these variables is heightened due to this replication.

References

Bryant, L. T., A Path Analysis Model for Secondary Physics Enrollments, Doctoral Dissertation, State University of New York at Buffalo February, (1974)

Bryant, L. T., A Path Analysis Model for Secondary Physics Enrollments, Journal of Research in Science Teaching, 14(3):177-189 (1979)

Chandavarkar, M., The Teaching and Learning of Physics in the United States, Doctoral Dissertation, Teachers College, Columbia University, New York (1988)

Darmondy, J. Development of Composite Background Variables, Second IEA Science Study, IEA/SISS bulletin 121 (1984)

Franz, J. The Crisis in High School Physics Teaching: Paths to a solution, Physics Today v36 n9 September (1983)

Joreskog, J. and Sorbom, D. LISREL VI - Analysis of Linear Structural Relationships by the Method of Maximum Likelihood, Morresville, Indiana, Scientific Software (1984)

Neuschatz, M., Covalt, M. 1986-87 Nationwide Survey of Secondary School Teachers of Physics, American Institute of Physics, New York (1988)

Noonan, R., Wold, H. Evaluation in Education: An International Review Series Evaluating School Systems Using Partial Least Squares, Volume 7, Number 3; Pergamon Press, New York (1983)

Weiss, I.R. Report of the 1985-1986 National Survey of Science and Mathematics Education, Research Triangle Institute, North Carolina, (1987)

Welch, Wayne W. Some Characteristics of High School Physics Students: Circa 1968 Journal of Research in Science Teaching. 6: 242-247; 1969.